

Data Abstraction and Integration of Information for Display: Analysis of Pilot-Automation Interaction

Asaf Degani & Michael Shafto
MS 269-4 NASA Ames Research Center
Moffett Field, CA 94035-1000, USA
{adegani, mshafto}@mail.arc.nasa.gov

Leonard Olson
825 Las Gallinas Ave. No. 308
San Rafael CA 94903
tetradtransformations@yahoo.com

ABSTRACT

This paper discusses some general challenges of information presentation and then presents a statistical approach for abstracting large amounts of data as well as a new theoretical approach for integrating the resultant information for the purpose of display. The example is from a study of crew interaction with the automatic flight control system of the Boeing 757/767 aircraft. We observed 60 flights and recorded every change in the aircraft control modes, as well as every observable change in the operational environment. The complete dataset consisted of 1665 such snapshots, each characterized by 75 variables. To quantify the relationships between the state of the operating environment and pilots' actions and responses, we used canonical correlation analysis because of its unique suitability for finding multiple patterns in large datasets. Traditionally, the results of canonical correlation analysis are presented by means of numerical tables, which are not conducive to recognizing multidimensional patterns in the dataset. Such patterns, however, are extremely important in characterizing the most important environmental conditions and their effects, and in revealing deviations (outliers) indicative of operational errors. We created a sun-ray-like diagram (which we call a heliograph) where all the independent variables are on the right side of the circle, and all the dependent variables are on the left. But since there are multiple patterns in the dataset, we employed Alexander's theory of centers to integrate all patterns into a single display. The theory describes 15 heuristic properties that help create wholeness in a design, and can be extended to the problem of data integration. Our ongoing work is to extend this theory to deal with problems of information abstraction and integration as well as packing large amounts of data for visualization.

Keywords

Integration of information, data abstraction, interface design, information systems, visualization.

INTRODUCTION

The complex systems of modern aircraft provide considerable amounts of data and information. We have onboard sensors, probes, and a multitude of data passing through data busses as well as data from onboard computers (e.g., FMS, autopilots, electronic checklists and alerting systems). In the future, with the introduction of Integrated Vehicle Health Monitoring (IVHM) technology, there will be even wider sensor coverage

available as well as another layer of nominal (expected) values that are computed for each sensor. In addition to systems data, it is becoming more and more feasible to provide extensive amount of other data (weather, ATC, navigational data, communication, dispatch, passenger, etc.) to the aircraft and make it available for both cockpit and cabin crews.

Nevertheless, the question of how to provide this wealth of information, so as to aid monitoring and help pilots in decision-making, is a tough one to answer. Given the limited display "real estate" and the limits on pilots' perceptual and cognitive resources, it is foreseeable that cockpit interfaces will eventually become the bottle neck of information. We foresee critical situations, where information about an abnormality exists, yet due to abstraction and display consideration this critical information is inaccessible to the crew. Since the amount of information available for display and use is only bound to increase in the future, we believe that such questions of data extraction and presentation deserve a thorough and fundamental consideration.

Our general approach to these issues centers on the ability to (1) extract continuous, discrete, and pattern-like data (from onboard sensors, for example), (2) organize it into models and databases, (3) abstract it, (4) integrate it in a meaningful way on the display, and finally (5) address the potential problems of ambiguity and uncertainty of the displayed information (e.g., failed sensors) as well as the problem of contingency procedures (e.g., diagnosis due to incomplete information) and recovery sequences.

As it stands today, the human-factors, applied psychology, and human-computer interaction (HCI) communities don't have a general approach and an established theory and methodology for assimilating, integrating, and packing large amounts of data so as to enhance and augment display content to make it more useful and effective. This thorny issue concerns not only information management and monitoring, but also interpreting and interacting with current and future automated control systems. We view the problem of human-automation interaction as a subset of the problem of information management.

Below is an example of data abstraction and information integration for the purpose of identifying patterns of pilot automation interaction and identification of anomalies (i.e., deviations from the norm that can potentially lead to an incident or accident). While the data and abstraction methods described here are intended for off-line monitoring and analysis, the principles of abstraction and integration are applicable also to on-line monitoring. In the following sections we describe the data, representation of statistical

patterns, and an approach for methods for integration and packing of multiple patterns into a single display. We conclude with several observations on data abstraction and information integration.

DATA STREAM OF PILOT-AUTOMATION INTERACTION

Canonical correlation analysis is a type of multivariate linear statistical analysis, first described by Hotelling [4]. It is used in a wide range of disciplines (such as chemistry, meteorology, and artificial intelligence) to analyze the relationships between multiple independent and dependent variables. The information presented in Figure 1 is derived from a canonical correlation analysis of a study of crew interaction with the automatic flight control system of the Boeing 757/767 aircraft. We observed 60 flights and recorded every visible change in the aircraft control modes, either manually initiated (e.g., the pilot selected a new mode) or automatically initiated (e.g., an automatic mode transition), along with all the settings relating to the flight control system status (e.g., waypoints and altitude values selected by the pilot). Likewise, every visible change in the operational environment (e.g., a new instruction from Air Traffic Control, or switching from one Air Traffic Control facility to another) was recorded, along with related variables such as the aircraft altitude, speed, and distance from the airport. In a way, it was like taking a snapshot of every change that took place both in and outside the cockpit. Overall, the dataset consisted of 1665 such snapshots, each characterized by 75 variables. Approximately half of the variables had to do with the operational environment and the other half had to do with pilot's responses [2].

In general, we were interested in identifying the relationships that exist between the state of the operating environment (independent variables) and pilots' actions and responses as represented through their interaction with the automatic flight control system and its modes and settings (dependent variables). The value of using canonical correlation in this case derived from its unique suitability for finding independent pairs of correlated patterns in large datasets.

REPRESENTATION OF STATISTICAL PATTERNS

Traditionally, the results of canonical correlation analysis are presented by means of numerical tables. However, a tabular format hinders the eye from recognizing and understanding the multidimensional patterns that exist in the data. Yet these patterns are extremely important, not only because they help the analyst characterize the most important environmental conditions and their corresponding effects on pilots' actions, but also because this method can reveal singular deviations from a well-established pattern (which may be indicative of an operational error that can potentially lead to an incident or accident). Using structured correlations (the correlations of the X canonical variate with each of the original independent variables, and of the Y canonical variate with each of the original dependent variables), but seeking to avoid tabular representation of the data, we created a sun-ray-like diagram where all the independent variables (X1, X2, ...) are on the

right side of the circle, and all the dependent variables (Y1, Y2, ...) are on the left. We chose a circle with rays to emphasize that "all variables are equal" (whereas employing a vertical and/or horizontal layout implicitly suggests some ordering). We call such a diagram a heliograph [5].

The canonical correlation analysis identified three sets of patterns that were operationally meaningful, statistically significant ($r = 0.95, 0.88, 0.72$; $p < 0.001$), and independent (orthogonal) of each other. Each one of the three sets contains two patterns—one positive and one negative—depicted by dark and white bars respectively. For example, concerning the outer ring ($r = 0.95$) the positive pattern (dark bars) indicates that for all independent variables (X's) *when*

- altitude is high (above the average of 13,000 feet),
- the phase of flight is "descent,"
- the Air Traffic Control facility is "approach control,"
- and the vertical clearance is "descend to altitude."

Then the corresponding modes and settings selected by the pilots are most likely to be:

- autopilot "engaged,"
- pitch mode in "flight level change,"
- and thrust mode in "cruise."

The reciprocal pattern (white bars) indicates that *when*

- the Air Traffic Control facility is "departure control,"
- and the vertical clearance is "climb to altitude."

Then the most likely mode/settings selected by pilots will be:

- autothrottles "engaged,"
- pitch mode in "vertical navigation."

[Note that the patterns not only identify which modes and settings are used ("engaged"), but also which modes and settings are *not* used. So with respect to the second pattern (white bars), we know that while being controlled by "departure control," pilots hardly ever use the autopilot (i.e., they are hand-flying the aircraft) and are not selecting any lateral guidance from the automatic flight control system. Such information has considerable operational importance for safety and training purposes.]

INTEGRATION AND PACKING OF MULTIPLE PATTERNS

The above-mentioned $r = 0.95$ set is only one of three sets of patterns identified by the canonical correlation analysis. And while it is possible to present each set separately, we decided to combine all sets within a single display in order to see the overall "story" of how the patterns relate to one another and cover the range of all possible variables (both X's and Y's). In order to create a composite figure from all six (or more) patterns we decided to use the properties, operators, and processes described in Alexander's theory of centers [1]. We are finding this theory, which was conceived in the field of architecture, to be extremely helpful and applicable for information presentation. In our ongoing work we seek to extend this theory to deal with problems of information integration and packing of large amounts of data.

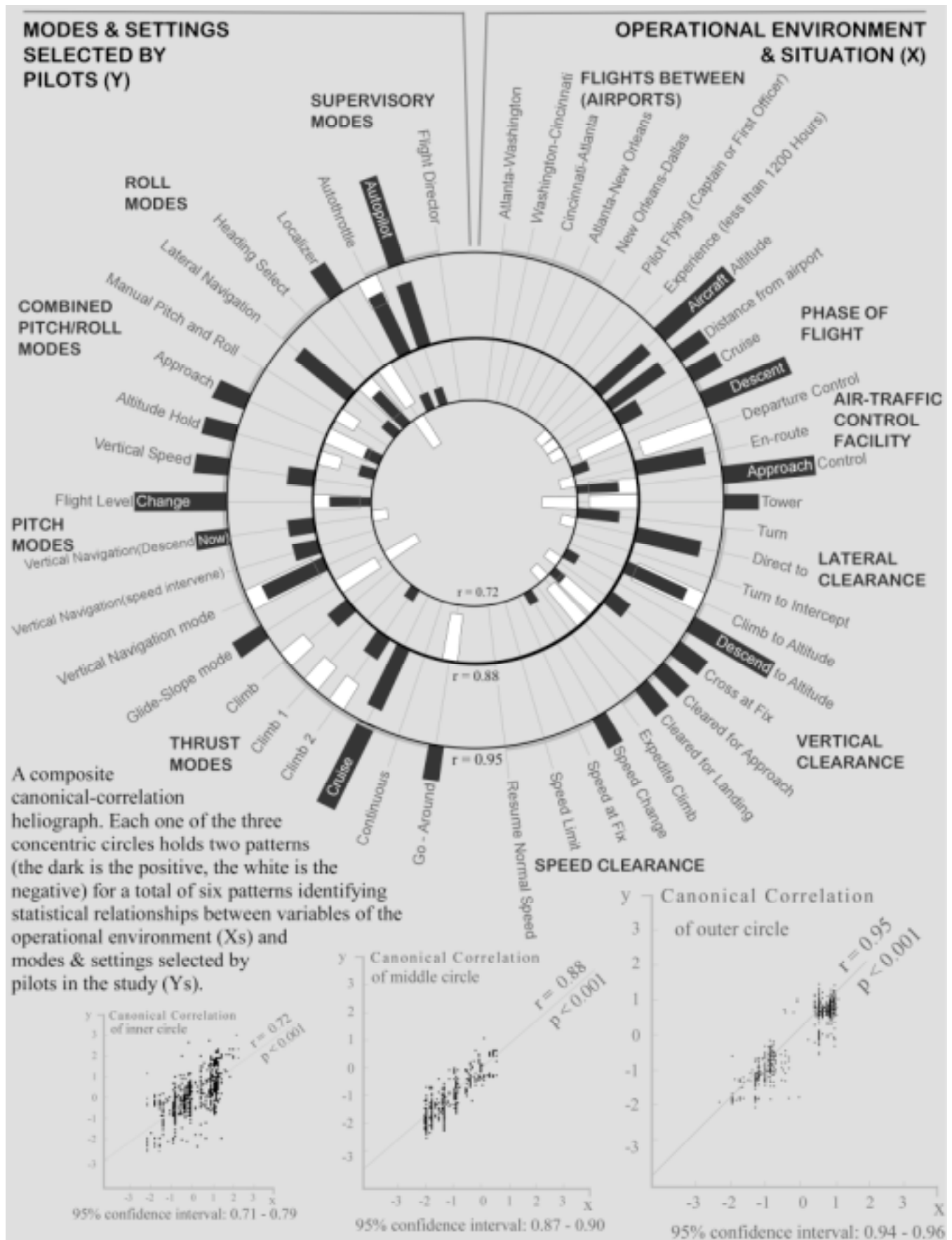


Figure 1. A composite diagram integrating three heliographs (and their bi-variate correlation plates).

Alexander's theory describes 15 heuristic properties that help create wholeness in a design, or a diagram, and which, for the purpose of our ongoing research, can thereby be extended and applied to the problem of data integration. The first property, *level of scale*, concerns the different ranges of sizes and internal coherence of "centers" within a given design. Thus, after realizing that there were several different levels of statistical strength (significance) among the three sets (0.95, 0.88, and 0.72), it became geometrically advantageous to pack them as concentric rings according to their statistical strength. (Note also the arrangement of the corresponding bi-variate correlation plates across the bottom of Figure. 1).

Alexander also describes a relationship between "centers" and "boundaries," showing that inviting and comprehensive living centers are often formed and strengthened by boundaries which tend to focus attention on the center and tie it with the surrounding space. (Just like a marsh serves as a boundary of the lake and a colonnade marks the end of a building and the beginning of the garden or street). In Figure. 1, the variable labels form a boundary between the inner world of data (values, significance, etc.) and the outer operational world. *Alternating patterns* and *echoes* are two other properties present in the ray-like spokes that guide the reader's eye as the rings (and variables) become smaller and merge into the center.

The center of the Figure forms a *void*, a profound property that is usually placed in the geometrical center of a design to draw the eye inward (e.g., the altar in a church or the empty space at the center of a mosque). We purposefully ordered the rings to (implicitly) suggest that as statistical significance decreases, the shrinking rings collapse into the void. Other utilized properties include *contrast* (between black and white bars), *interlock* (the overlap, and therefore the contrasting significance) between black and white bars of the same variable) and *gradients* (in the magnitude of bar sizes, which, for the purpose of this display, was abstracted into three categories—strong, weak, and none).

The properties used to create the Figure act together to create a literal sense of wholeness. This allows the reader to inspect the sum total of the patterns in this dataset and identify regions where there is intensity of coverage (where bars of a certain cluster are juxtaposed and where interlocks exist along a certain variable axis), as well as regions on the circumference of the circle that are empty—indicating variables, mostly on the environmental (X) side, that are not important and do not contribute much to pilots' responses. For example, the fact that the "flights between airports" is not important provides a meaningful piece of the puzzle: It assures us, as the analysts, that there is nothing of major importance about the idiosyncrasies of particular flights. In other words, the patterns are consistent over different flight legs—an important fact about their generality.

OBSERVATIONS AND FUTURE RESEARCH

In this paper, we have illustrated a process of data abstraction and information integration based on analysis of flightcrew interaction with automation. We believe that these concepts and methods are key aspects for monitoring, analyzing, and

interacting with data-rich environments such as networks, vehicle health monitoring systems, aircraft and spacecraft systems. Given the limited display "real estate," the (fixed) users' perceptual and cognitive resources, and the exponential growth in data availability, it is clear that visual displays and human computer interfaces represent a limitation which will only become more severe in the future.

Specifically, we focused our attention on two aspects of information presentation: abstraction of data and integration of information. The abstraction in the above example is statistical in nature with special emphasis on patterns. Other forms of statistical abstraction such as summary statistics (mean and standard deviation, range, mode, to name a few) as well as more analytical statistics (correlations, analysis of variance and more) are available. Nevertheless statistical abstraction is only one of many other abstraction methods. Reduction is yet another method of abstraction, where inconsequential data points or states are removed and abstracted away. Other reduction methods involve aggregation of states and data point by encapsulating them into sets that are then represented as a single entity (see [3] for an example of how a reduction method can be used to generate succinct and correct interfaces).

While the topic of abstraction of statistical data with all its variants (summary, analytical, patterns) has been around for quite some time and has a well defined and mathematically-based theory behind it, the topic of information integration and its application to packing of data and compression of visual information is in its infancy. While the underlying concepts of "order," "wholeness," and "coherence" are discussed by artists and architects — these terms are loosely defined and have multiple meanings depending on the speaker and his or her background. Currently, there does not exist a well defined theory for information integration let alone any mathematical foundation for its consistent application. We believe that this is a serious shortcoming that is worthy of further research where concepts form art, functional design, and architecture (e.g., Alexander's theory of centers) are brought in, extended to information presentation, and formalized. Our ongoing research is in developing a theory and formal methods for generating abstracted and integrated interfaces.

REFERENCES

1. Alexander, C. (2002). *The Phenomenon of Life*. Berkeley, CA: The Center for Environmental Structure.
2. Degani, A. (1996). *Modeling human-machine systems: On modes, error, and patterns of interaction*. Unpublished doctoral dissertation. Atlanta, GA: Georgia Institute of Technology.
3. Heymann, M., Degani, A. (Accepted). Formal analysis and automatic generation of user interfaces: Approach, methodology, and an algorithm. *Human Factors* (paper accepted for publication).
4. Hotelling, H. (1935). The most predictable criterion. *Journal of Educational Psychology* 26, 139-142.
5. Shafto, M., Degani, A., Kirlik, A. (1997). Canonical correlation analysis of data on human-automation interaction. *Proceedings of the 41st Annual Meeting of the Human Factors and Ergonomics Society*. Albuquerque, NM

